DOCUMENT RESUME

ED 103 809 CS 001 667

AUTHOR Adams, Charles F.
TITLE Effect of Comprehensive Achievement Monitoring on Student Learning.
PUB DATE Nov 74
NOTE 52p.; Presented as a paper at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS *Computer Assisted Instruction; Elementary Education; Reading; *Reading Achievement; *Reading Development; Reading Improvement; Reading Programs; *Reading Research; *Reading Skills

ABSTRACT

Three years of data on student achievement in reading were collected and analyzed to provide an empirical base to test the two hypotheses under investigation: there is no significant difference in the increase in student achievement between those students involved in a program using a comprehensive achievement monitoring (CAM) system and those students not using a CAM system; and there is no significant difference in the achievement levels of students as measured by either criterion referenced tests (CAM) or standardized norm referenced tests (Metropolitan Achievement Test). CAM, a computer supported pupil and program evaluation system based upon a criterion referenced model of evaluation, focuses on students' achievement of specific learning outcomes rather than on global educational outcomes. All the fourth and sixth graders receiving regular instruction in reading were chosen to provide the sample population. The results indicated that there were no significant differences between the experimental and control groups on the Metropolitan Achievement Test. In posttest comparisons between the student groups using CAM and those not using CAM, the student achievement of reading skills objectives was significantly higher for those using CAM. (WR)

ED103809

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EFFECT OF COMPREHENSIVE ACHIEVEMENT MONITORING ON STUDENT LEARNING

Prepared by: Charles F. Adams
November, 1974

CS 001 667

PREFACE

To conduct a research study in public education requires the cooperation and effort of a large number of people. Especially important are the attitude and willingness of school administrators to support and contribute to research efforts that generally require time and resources beyond normal operations. The contribution of people in the West Seneca Central School District to the study have reported is worthy of commendation.

There were so many people, teachers, clerical personnel and especially the students themselves of the West Seneca District who contributed to the research effort that to single out only a few seems arbitrary. Never-the-less, I must hazzard that risk and mention a few individuals without whose cooperation and effort the study could not have taken place. First, Dr. Carl Markello, Deputy Superintendent of Schools, whose leadership over the three years was extremely supportive. Rarely does one find a school administrator with the knowledge and passion for research as Dr. Markello. Both the principal and assistant principal of the Northwood Elementary School (Mr. Al Wolchuck and Mr. Steven Maricich, respectively) cooperated and supported the study in inumerable ways. Two of the finest teachers I have had the occasion to work with, Ms. Mary Alice Walz and Ms. Ann Wojiechowich, demonstrated that quality of dedication to the profession that is often overlooked these days. Without their hard work and patience with the researcher, the study could not have been completed. Finally, a special commendation for the skill and patience of Ms. Doretta Dodge for her assistance in the tedious task of data tabulation and analysis.

All of these, both mentioned by name and those un-named, have my sincere thanks for all the effort expended over the past three years.

CFA

I. Introduction

In the school year 1971-72 the Erie County Board of Cooperative Educational Services #1 (BOCES) offered a new criterion referenced curriculum development and evaluation service to local school districts on a pilot basis. This service is called the System for Pupil and Program Evaluation and Development (SPPED) which is a developmental project sponsored by the Bureau of School and Cultural Research of the New York State Education Department. The primary component of the SPPED that was implemented on a pilot basis was the Comprehensive Achievement Monitoring (CAM) evaluation system.

CAM was originally initiated under a grant from the Charles F. Kettering Foundation to Dwight W. Allen at Stanford University in 1967. Primary developers of the system were William P. Gorth and Paul Pinsky. In 1968 the project moved to the University of Massachusetts where it came to fruition. CAM has undergone further refinements in New York State under the direction of Robert O'Reilly, Chief, Bureau of School and Cultural Research.

Comprehensive Achievement Monitoring is a computer supported pupil and program evaluation system based upon a criterion referenced model of evaluation. CAM focuses on students achievement of specific learning outcomes (behaviorally stated objectives) rather than on global educational outcomes. CAM requires the specification of a complete set of objectives for a course and development of specific criterion referenced test items to measure each objective. Thus, CAM focuses on the actual achievement of specific behavioral objectives rather than relative (or norm-referenced) measures of achievement.

The evaluation design underlying the CAM approach is quite different from the more traditional approach to student and program evaluation. Each course objective is tested frequently throughout the duration of the course (semester or school year).

> At each test administration, performance on objectives not yet taught is pretested, performance on objectives just taught is immediately post-tested, and performance on objectives taught earlier in the course is measured for retention. Parallel test forms, comparable in difficulty and content, are all used at each test administration, but each student receives a particular (test) form only once during the course. Each form typically has an item for each objective. Each item is used on only one test form. The function of a particular item changes in relation to the time at which the objective it is measuring is taught. Testing may take place at regular intervals (e.g., every two weeks) or at the end of certain instructional units.[1]

Due to the frequency of testing, the large amounts of data analyzed, and the numerous types of decision-relevant reports possible, computerized data processing is the only efficient and effective system to employ. Output from CAM Computer programs at each test administration provides the following information:

1. Individual Student Data - a total score for that test and each previous one. A total score on the objectives for which the student has received instruction for that test and each previous one. A correct-incorrect indication for each item on the test coded to the specific course objective it measures.
2. Group Summary Data - for each group or subgroup of students (e.g., class, ability group, grade level, etc.); the percentage correct from all test items on all forms for each objective or groupings of objectives such as modules or units.
3. Item Analysis Data - periodically as desired, but usually at the end of a course, data on each item is provided. The information treats each item by its three functions, pretest, posttest, retention and provides data on its difficulty level and the distribution of choices for multiple choice distractors.

---

[1] Gorth, William P. "Comprehensive Achievement Monitoring (CAM): A Project to Develop Longitudinal Classroom Evaluation Using Item Sampling." Paper presented at National Council of Measurement of Education. New York, February, 1971, p. 7.

Comprehensive Achievement Monitoring, then, is a systematic way of measuring student achievement of specific learning outcomes employing a longitudinal testing design and the use of computer programs for rapid data analysis and reporting of the results to students, teachers and others. In addition to providing specific information on individual students, the summary data is useful for program evaluation.[2]

II. Research Problem

During the first year in which the BOCES offered the CAM evaluation component of SPPED to a few selected school districts, one of the pilot districts, West Seneca, agreed to conduct a research study cooperatively with the BOCES. Both the local school district and the BOCES Instructional Services staff were interested in determining the impact this technological innovation would have upon the school program and student achievement. The local school district personnel were primarily concerned with whether or not the utilization of CAM would improve student achievement. Both the pilot district and BOCES were hopeful that student achievement in reading would be positively affected by employing this computer based evaluation system. The basic premise underlying the employment of the CAM system was that the receipt of CAM test results on a regular basis would provide students and teachers with relevant data at appropriate times that would increase the effectiveness of the teaching-learning process. The BOCES researcher was also interested in CAM's impact upon student achievement but also wished to answer the question, "Which form of program evaluation, Norm-referenced (NRT) or Criterion-referenced,(CRT), is a more sensitive measure of student learning outcomes?".

---

2 For a more detailed explanation of the CAM system see:
Gorsh, William, O'Reilly, Robert and Pinsky, Paul, Comprehensive Achievement Monitoring. Amherst, Massachusetts; University of Massachusetts, 1974.

The pilot school district and the BOCES researcher agreed that the proposed study would address itself to both concerns.

Evaluation specialists as well as public school personnel have been examining and debating the CRT vs NRT question for a number of years.[3] The general contention of adherents of CRT is that since "A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards",[4] they are more appropriate measures of the intended outcomes of an instructional program. The use of CAM criterion referenced tests as well as a standardized norm referenced measure to assess program outcomes in the pilot school district provided an opportunity to investigate this contention.

Thus, there were two basic propositions under investigation:

1. The use of the CAM evaluation system will significantly increase student achievement.

2. Criterion referenced measures used in the CAM system are more sensitive in detecting instructional program outcomes than standardized norm-referenced measures.

Reformulated into research hypothesis:

H_1: There is no significant difference in the increase in student achievement between those students involved in a program using the CAM system than those students not using a CAM system.

H_2: There is no significant difference in the achievement levels of students as measured by either criterion referenced tests (CAM) or standardized norm referenced tests (Metropolitan Achievement Test).

---

3 See Cronlack, L.J. "Course Improvement Through Evaluation" Teachers College Record, 1963, 43, 672-683 and Barnabei, Raymond and Leles, Sam. Behavioral Objectives in Curriculum and Evaluation. Dufuque: Kendall/Hunt Publishing Company. Ross, Paul C. "Some Considerations in the Design and Use of Criterion-Referenced Tests". Paper presented at Northeast Educational Research Association, March, 1970.

4 Robert Glazer and Anthony J. Nitko, "Measurement in Learning", in Educational Measurement, edited by Robert L. Thorndike, Washington, D.C. American Council on Education, 1971, p. 653.

III. Research Design and Methodology

In conducting research studies in Education the operational circumstances of public schools often preclude the establishment of rigorous research designs prior to undertaking a specific investigation. Such was the case in this study. Though the local school district, West Seneca, was quite cooperative throughout the entire study, much of the data for the study was based upon that which was available via normal school operations. Also, much of the design could be considered post hoc formulation since the specification of the design evolved over the three year period of the study.

At the commencement of the study, approximately December of 1971, the West Seneca Central School District had been utilizing the CAM evaluation system in the intermediate reading program (grades 4-6) in one elementary school building since the beginning of the 1971-72 school year. During the previous summer five teachers from the school district attended a four week workshop where they developed the instructional objectives and criterion referenced test items for their reading program. Therefore, the program had been in operation for a few months prior to the decision to conduct the study.

The initial research design of the study was an "experimental-control" design employing pre and post assessment of student achievement in reading. The school district selected two elementary schools, the Northwood and Clinton Elementary Schools, to serve as the sample populations for the study. Both schools serve a student population residing in the same geographical area. The district was of the opinion that the background of the students, in terms of socio-economic status, wealth, ethnic origins, etc. was very similar for both school populations. The Northwood School which was using the CAM system was the experimental group and the Clinton School was the control group.

In this first year of the study (the extension of the study into a three year longitudinal design was determined after the results were reviewed at the end of year one) two groups of students in each school were selected as the subjects for the study. All the fourth and sixth grade students receiving regular instruction in reading were chosen to provide the following sample populations:

	Experimental (Northwood)	Control (Clinton)	Total Students
4TH Grade	154	139	293
6TH Grade	131	89	210

Since the effect of CAM on student achievement in the reading program was the major variable under study, controls for as many other variables as possible were attempted. The characteristics of the teachers in both schools (age, experience and training) were examined. The teaching populations in both schools in terms of these variables were very similar. The curriculum and instructional resources (basal readers) were the same for both schools. The policy on student grouping for instruction was the same for both schools. The only major difference between the two groups was that some of the teachers in the experimental group received four weeks of training during a summer workshop on establishing and utilizing a CAM evaluation system.

Probably the most important variable for which controls were needed was the achievement level in reading of both student groups at the beginning of the school term. Shortly after the beginning of the 1971-72 school year, the standardized Metropolitan Achievement Tests (MAT) were administered to both groups of students. The results of the reading section of these tests are reported in Tables 1 and 2.

TABLE 1

Comparison of Mean Standard Scores on the Pre-MAT - Grade 4

	Experimental Group	Control Group
Mean Standard Score	65.9	63.0
Standard Deviation	11.6	13.1
Sample Population Size	156	152

TABLE 2

Comparison of Mean Standard Scores on the Pre-MAT - Grade 6

	Experimental Group	Control Group
Mean Standard Score	83.4	84.3
Standard Deviation	16.1	12.7
Sample Population Size	133	98

The experimental group at the fourth grade level has a statistically significant higher mean score on the pre-test than the control group. The calculated statistic was 2.0567 which was significant at the $P < .05$ level of error. At the sixth grade level the reading ability of the students, as measured by the MAT, was almost the same, there being only .9 difference in the mean standard scores of the two groups. A T test indicated no significant difference. Thus, the sixth graders in the two groups began the school year at approximately the same achievement level, but the fourth graders in the experimental group began the year with a higher achievement level than the control group. Since the school year was well underway when the research study began, it was not possible to obtain pre instruction measures with the CAM tests for the control group.

Teachers from the experimental group developed a set of reading objectives during a summer training session. In addition to a set of instructional objectives for both fourth and sixth grade reading courses the teachers also developed a set of criterion referenced test items to measure the objectives. At the beginning of the school year the fourth and sixth grade teachers from the control group were presented with the set of reading objectives and were asked to indicate which objectives they planned to utilize in their reading program. In all cases, the teachers of the control group indicated that over 95% of the objectives were ones that they attempt to teach toward in their program. Thus, both the experimental and control groups were similar (with the exception of higher achievement level at the 4th grade by the experimental group) and both were being taught the same basic reading program with the same instructional resources. The major difference between the two groups was the utilization of a CAM evaluation system by the experimental group. Eight times a year the experimental student groups were given CAM criterion referenced tests and the students and their teachers received analysis reports on the results.

The final (eighth) CAM test was administered at the end of the school year to both the experimental and control groups of students. The scores on these tests were used as the post instruction criterion referenced measure of student achievement in reading, as were the scores on the norm referenced standardized Metropolitan Achievement Tests in Reading which were also administered in June of 1972. Mean scores for both tests for each group were computed and tests were employed to determine if there were any significant differences in student achievement.

After a review of the first year findings by the school district administration and the BOCES researcher, it was decided that the study should be continued for at least another year.

Due to organizational arrangements and the procedure for the assignment of students to classes, all of the students who participated in the first year of the project were not involved in the second year. The sixth grade students in year one were dispersed into middle school programs which were not using CAM, thus it was impossible to follow their progress in the second year. The fifth grade population in both the experimental (CAM) and control groups (fourth graders the previous year) were the populations under study in the second year. Table 3 depicts the populations involved in this study.

TABLE 3

	Experimental (Northwood)	Control (Clinton)	Total Students
5th Grade	145	129	274

As in the first year, two measurement approaches were utilized to assess student achievement in reading. The criterion referenced CAM tests developed by the 5th grade teachers of the experimental group and the norm referenced Metropolitan Achievement Test (MAT) in reading were administered to both student groups during the first week in June, 1973. The second year design employed only post tests. The fifth grade reading program employed the same basic CAM evaluation design as had the 4th and 6th grade programs the previous year. There were eight criterion referenced tests administered to the experimental group at regular intervals throughout the year. These CAM tests

had been revised, based upon item analysis data provided during the first year of the project. It was the result of the item revision process that gave rise to the second hypothesis under investigation in this study. The teachers of the experimental group having had a year's experience working with the criterion measures, were able to refine the criterion test items to more accurately assess student achievement of the program objectives. This assumption along with the inconclusive results in year one led the researcher to believe that the criterion measures would become more sensative to the effects of a program over time.

In the third year of the study the same basic experimental-control design was used. However, the data collection was expanded to involve as many students as possible and to assess background information on both populations of students. The same two schools remained K-5 buildings as in year two and both the entire fourth and fifth grade students in both schools were chosen as the sample population. The fifth grade students in the experimental group were involved in the CAM program the previous year. Table 4 depicts the populations involved.

TABLE 4

CAM Study Populations - Year 3

	Experimental (Northwood)	Control (Clinton)	Total
4th Grade	134	157	291
5th Grade	141	141	282
			573

To empirically validate the assumption that the family background of both populations of students were similar, a questionnaire was sent to the parents of each fourth and fifth grade student enrolled in both schools. (See Appendix C for copy of questionnaire). Five hundred and ninety-two questionnaires were mailed with four hundred and sixty-seven returned for a return percentage of 78%.

Responses to the survey questionnaire are summarized in Tables 5 and 6.

Table 5

Student Background Data - 4th Grade

Question 1: Father's education														
Responses	College Degree		Some College		H.S. Graduate		Some High School		8th Grade Graduate		Less Than 8th Grade		No Response	
	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%
Experimental N=142	24	17%	19	13%	67	47%	27	19%	2	1%	1	1%	2	1%
Control N=128	15	12%	34	27%	52	41%	20	15%	2	2%	2	2%	3	2%

Key: Fr. = Frequency
% = Percent

Question 2: Mother's education												
Responses	Attended College		High School Graduate		Some High School		8th Grade Graduate		Less Than 8th Grade		No Response	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Experimental N=141	25	18%	84	60%	23	16%	4	3%	1	1%	4	3%
Control N=127	25	20%	81	64%	15	12%	4	3%	1	1%	1	1%

Key: Freq. = Frequency
% = Percent

Question 3: Is mother working?				
Response	Yes		No	
	Frequency	%	Frequency	%
Experimental	40	29%	100	71%
Control	41	32%	86	68%

* Percentages are rounded to nearest percent.

Table 5
(Continued)

Question 4: Number of siblings														
Responses	None		One		Two		Three		Four		Five		Over Five	
	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%
Experimental N = 142	3	2%	26	18%	59	42%	33	23%	13	9%	6	4%	2	1%
Control N = 127	7	6%	34	27%	36	28%	26	20%	11	9%	10	8%	3	2%

Key: Fr. = Frequency
% = Percent

Question 5: Number of other schools attended												
Responses	No other		One		Two		Three		More than Three		No Response	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Experimental N = 142	61	43%	62	44%	13	9%	2	1%	1	1%	3	2%
Control N = 127	86	68%	32	25%	8	6%	0		0		1	1%

Key: Freq. = Frequency
% = Percent

Question 6: Was student's attidute toward reading changed positively?								
Responses	Yes		No		No Change		No Response	
	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
Experimental N = 141	105	74%	12	9%	23	16%	1	1%
Control N = 127	65	51%	16	13%	44	35%	2	2%

Table 6

Student Background Data - 5th Grade

Question 1: Father's Education

Responses	College Degree		Some College		H.S. Graduate		Some High School		8th Grade Graduate		Less than 8th Grade		No Response	
	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%
Experimental N = 101	16	16%	20	20%	40	40%	18	18%	5	5%	1	1%	1	1%
Control N = 101	20	20%	19	19%	39	39%	16	16%	4	4%	2	2%	1	1%

Key: Fr. = Frequency
% = Percent

Question 2: Mother's Education

Responses	Attended College		H. School Graduate		Some High School		8th Grade Graduate		Less than 8th Grade		No Response	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Experimental N = 99	13	13%	62	62%	20	20%	3	3%	0	---	1	1%
Control N = 117	18	17%	60	58%	21	20%	4	4%	0	---	---	---

Key: Freq. = Frequency
% = Percent

Question 3: Is mother working?

Responses	Yes		No	
	Frequency	Percent	Frequency	Percent
Experimental N = 98	31	32%	67	68%
Control N = 101	27	27%	74	79%

Table 6
(Continued)

Question 4: Number of siblings														
Responses	None		One		Two		Three		Four		Five		Over Five	
	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%	Fr.	%
Experimental N = 105	4	4%	28	27%	40	38%	15	14%	11	10%	7	7%	0	0
Control N = 101	2	2%	19	19%	30	30%	23	23%	20	20%	4	4%	3	3%

Key: Fr. = Frequency
% = Percent

Question 5: Number of other schools attended												
Responses	No other		One		Two		Three		More than Three		No Response	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Experimental N = 99	47	47%	42	42%	9	9%	1	1%	0	0	---	---
Control N = 102	58	57%	34	33%	7	7%	1	1%	1	1%	1	1%

Key: Freq. = Frequency
% = Percent

Question 6: Was student's attitude toward reading changed positively?								
Responses	Yes		No		No Change		No Response	
	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
Experimental N = 99	62	62%	13	13%	23	23%	1	1%
Control N = 102	59	58%	15	15%	27	26%	1	1%

Key: Freq. = Frequency

The family background of the students in both the experimental and control schools was found to be similar. There was a tendency toward slightly smaller families, in terms of the number of children, in the population served by the experimental school (Question 4). The families in the experimental school also are somewhat more mobile than those in the control school population as evidenced by the number of other schools attended (Question 5). Responses to question 5 will be discussed in Section V.

The same achievement measures were utilized in the third year as in the previous years, both the CAM tests and the MAT. Again the CAM tests had undergone another revision prior to the start of the school year. The same teachers as in previous years were responsible for the revisions based upon item analysis data received at the end of the school year (June 1973). In this year of the study all students were pre and post tested with both the appropriate level CAM and Metropolitan Achievement Tests. Thus, student achievement gains in reading were determined as measured by both normative and criterion referenced tests.

From the experience of the first two years of utilizing the CAM system, the teachers in the Northwood School (Experimental Group) determined that one set of objectives and criterion measures was insufficient for all students at any one grade level. During the second year of implementation two sets of objectives and criterion measures were developed for both the fourth and fifth grade reading programs. In effect there were four CAM courses covering the two year span from beginning fourth grade to advanced fifth grade. Thus a more continuous progress curriculum was established. This enabled the students to progress from one set of objectives (course) to another at their individaul learning rates.

In the third year of the study, there were a number of students instructed at both fourth grade levels. At the fifth grade level only one of the two levels of objectives was used for instruction. Thus, three sets of CAM criterion referenced tests were administered to the fourth and fifth grade in both schools on a pre and post basis.

IV. Data Analysis

For each of the three years in which achievement data were collected for both experimental and control groups of students mean scores on each of the four group measures were calculated. T tests were employed to determine statistical significance of differences between these mean scores. Tables displaying the results of the data analysis are contained in Appendix A. In addition, charts depicting criterion referenced score distributions of sub populations of students are found in Appendix B.

Comparisons of reading achievement between the experimental and control groups of students in the first year produced mixed, inclusive results. As was indicated earlier, the 4th grade students in the experimental group in year one had a statistically higher pre-instruction score on the MAT than the control group while there was no statistically significant difference between the two 6th grade groups. Thus, the 4th graders in the experimental group started out with a higher achievement level than the control group.

The results of the Metropolitan Achievement Tests (MAT) in reading which were administered to students in the experimental and control schools at the end of the year were similar to the pre-instruction results. The mean standard score for the fourth grade experimental group was found to be significantly higher than the control group when a T test of significance was employed. The T statistic was

Calculated as 2.0819 which was significant at the $P < .05$ level of error. Though no statistical test of significance was employed, the grade equivalency score for the experimental group was three and a half months higher than the control group. Though the control group had a mean standard score two points higher than the experimental group at the sixth grade level, it was not found to be statistically significantly higher. The T statistic computed has a value of -1.0526 which was not significant at any acceptable level of error. The control group had a slightly higher mean grade equivalent score than the experimental group.

The results of post-instruction CAM testing in year one indicated the same pattern as the MAT results. The CAM tests at the fourth grade level measured twenty-one different reading objectives with twenty-one different test items. The experimental group of fourth graders had a mean score of 16.41 items correct; whereas, the control group had a mean score of 13.31 items correct. A T test for statistical significance was employed producing a value of 7.56, indicating a significantly higher mean score for the experimental group at the $P < .01$ level of error. The sixth grade CAM tests consisted of seventeen items on each form measuring all seventeen reading course objectives. The experimental group had a mean score of 10.75. The T statistic computed resulted in a value of .223 which indicated there was no statistically significant difference between the mean scores at the sixth grade.

Further data analysis of the CAM test showed marked differences between the fourth grade experimental and control groups, when teacher sub-sections of each course were examined. The results are presented in Table 7.

TABLE 7

Mean Scores of Teacher Sub-Sections for CAM Tests - Grade 4

	Experimental Group	Control Group
Sub-section 1	13.96	12.15
Sub-section 2	14.12	12.86
Sub-section 3	16.73	13.37
Sub-section 4	17.10	13.82
Sub-section 5	19.34	14.23

Three of the class sections in the experimental group had mean scores of the CAM test higher than the highest class section of the control group. In one sub-section (mean score 19.34) ten students had a perfect score. Graphs of all student scores on the CAM test can be found in Appendix B.

Consistent with the overall comparison at the sixth grade level between the experimental and control group using criterion referenced measures is the comparison by class sub-section. The data indicate very little variation between groups.

TABLE 8

Mean Scores of Teacher Sub-sections for CAM Tests - Grade 6

	Experimental Group	Control Group
Sub-section 1	10.07	10.27
Sub-section 2	10.80	10.93
Sub-section 3	10.94	11.07
Sub-section 4	11.41	

In year two only the students who had been 4th graders the previous year (5th graders in year two) were included in the study. Also, only post-instruction measures were administered to both groups (experimental group was pre-tested with the CAM tests).

The CAM tests consisted of thirty-four items measuring 34 separate instructional objectives in reading. The experimental group had a mean score of 25.2 items correct; whereas, the control group has a mean score of 23.1 items correct. A T test for statistical significance was employed producing a T value of 3.0882, indicating a significantly higher mean score for the experimental group at a $P < .01$ level of error.

The results of the MAT in reading indicated a mean standard score of 83.0 for the experimental group and a mean standard score of 80.9 for the control group. The T statistic calculated was 1.4286 which was significant at the $P < .10$ level of error but not at the .05 level.

Although the actual differences in achievement were statistically significant the magnitude of the difference is not exceptionally large, especially the results of the standardized MAT. A score value of 2.1 difference in mean scores between groups for both types of measures was obtained. The variance of student scores on both tests for both groups is also similar. The distribution of scores on the CAM tests are presented in Charts 5 and 6 found in Appendix B. One can observe that student scores for the experimental group are skewed toward the maximum to a slightly greater extent than the scores of the control group students.

This same group of students tested in a similar manner the previous year had approximately the same results. The two year comparison of post-instruction mean scores is contained in Table 9.

TABLE 9

Two Year Comparison of Mean Scores on CAM and MAT

	EXPERIMENTAL GROUP		CONTROL GROUP	
	4th Grade	5th Grade	4th Grade	5th Grade
CAM Score	16.41	25.2	13.31	23.1
MAT Mean Standard Score	65.9	83.0	63.0	80.9
Sample Population Size	128	140	88	129

In year three of the study the curriculum had undergone a major revision. There were now four sets of course objectives and criterion measures for the two grade levels. The design was more rigorously employed in year three with pre and post assessments made to the experimental and control groups at both 4th and 5th grade levels. Three of the four course curricula and sets of CAM measures were used in the third year. Since three different sets of objectives (3 CAM courses) were employed during the year, comparisons between the experimental and control groups were made for all three courses.

A T statistic was computed to test the difference in grade equivalent mean scores on the MAT pre-tests for both fourth and fifth grade students. (Tables A-9 and A-10 Appendix A). The T values for the comparisons between experimental and control groups at the fourth grade level was .000 and at the fifth grade level was -.200, neither of which was statistically significant. Thus, there was no significant difference in reading achievement between experimental and control groups at the beginning of year three as measured by norm referenced tests for either fourth or fifth grade students.

The post-instruction comparisons of achievement measured by the MAT for fourth and fifth grade students are reported in Tables 10 and 11.

TABLE 10

Comparison of Post-test Mean Scores on MAT - Grade 4

	Experimental Group	Control Group
Mean Standard Score	73.21	72.04
Mean Grade Equivalent Score	5.0	4.3
G.E. Standard Deviation	1.6	1.7
Sample Population Size	132	157

TABLE 11

Comparison of Post-test Mean Scores on MAT - Grade 5

	Experimental Group	Control Group
Mean Standard Score	81.24	80.66
Mean Grade Equivalent Score	6.1	6.0
G.E. Standard Deviation	1.9	1.8
Sample Population Size	136	141

The calculation of the T statistic for both fourth and fifth comparisons of G.E. mean scores produced the following respective values, 4th = 3.50 and 5th =.223. At the fourth grade level the T value was statistically significant at the P< .01 level of error. Thus the experimental group had a higher reading achievement level as measured by the Metropolitan Achievement Test at the end of the school year, although the mean standard score was only 1.17 points higher. At the fifth grade level there was not a statistically significant difference in the post-test mean grade equivalent scores between the experimental and control group.

The following set of tables, 12, 13, and 14 display the results of the analyses of CAM test data for year three. Again, both pre and post comparisons are made between the experimental and control groups for each of the three CAM courses.

TABLE 12

Comparison of Pre-test and Post-test CAM Mean Scores

Grade 4 - (Course 409)

	Experimental Group	Control Group
Pre-test Mean Score	10.3	10.56
Pre-test Standard Deviation	4.8	3.5
Sample Population Size	40	44
Post-test Mean Score	19.3	10.2
Post-test Standard Deviation	3.7	3.9
Sample Population Size	42	19

The CAM tests for course 409 had 24 test items on each form measuring twenty-four different course objectives. At the beginning of the year, the pre-test mean scores for both groups was almost identical with the experimental group mean score of 10.3 and the control group mean score of 10.56. The T statistics calculated produced a value of -.280 which was not statistically significant. However, the post measure at the end of the year indicated a mean score of 19.3 for the experimental group and 10.2 for the control group. The calculated T value was 8.270 which is statistically significant at the P< .01 level of error. On the average, the students in the experimental group achieved nine more reading objectives during the year than the control group.

TABLE 13

Comparison of Pre-tests and Post-tests CAM Mean Scores - Grade 4 (Course 419)

	Experimental Group	Control Group
Pre-Test Mean Scores	13.0	15.16
Pre-Test Standard Deviation	4.9	4.6
Sample Population Size	115	137
Post-test Mean Scores	20.1	14.7
Post-test Standard Deviation	5.1	4.8
Sample Population Size	125	163

The CAM tests for course 419 had 27 items measuring 27 different course objectives on each of the eight test forms. The pre-test results indicate a mean score of 13.0 correct items for the experimental group and 15.16 items for the control group. The T statistic had a calculated value of -3.600 which was significant at the P< .01 level of error.

At the end of the year the post-test results indicated a mean score of 20.1 correct items for the experimental group and 14.7 items for the control group.

The value of the T statistic was 9.000 which was significant at the P< .01 level of error. Though the experimental group started the school year with a statistically significant lower achievement level in reading than the control group, as measured by the CAM tests, they achieved a significantly higher level of achievement at the end of the year. There was an average gain of over seven objectives by the experimental group while none at all for the control group.

TABLE 14

Comparison of Pre-test and Post-test CAM Mean Scores - Grade 5 (Course 509)

	Experimental Group	Control Group
Pre-test Mean Score	21.6	20.96
Pre-test Standard Deviation	5.4	5.9
Sample Population Size	112	111
Post-test Mean Score	25.6	22.7
Post-test Standard Deviation	5.4	5.6
Sample Population Size	108	114

There were thirty-seven items on the CAM test forms for course 509 measuring 37 different course objectives. The pre-test mean scores were 21.6 for the experimental group and 20.96 for the control group. The T statistic calculated had a value of .850 indicating no statistical difference in the mean scores between the groups. At the end of the school year the post-test mean scores were 25.6 for the experimental group and 22.7 for the control group. Employing a T test to determine whether the difference between mean scores was statistically significant, produced a T value of 3.920. This was statistically significant at the P< .01 level of error. The fifth grade students in the experimental group had gained significantly more than the control group.

The third year data was also analyzed by teacher sub-sections of the courses. Tables A-11, A-12, A-13 in Appendix A display the mean scores by teacher sub-sections for the three experimental groups. This analysis was not possible for the control grou

since data was not submitted by teacher sub-section.

V. Conclusions and Implications

Three years of data on student achievement in reading were collected and analyzed to provide an emperial base to test the hypotheses under investigation:

H_1: There is no significant difference in the increase in student achievement between those students involved in a program using a CAM system and those students not using a CAM system.

H_2: There is no significant difference in the achievement levels of students as measured by either criterion referenced tests (CAM) or standardized norm referenced tests (Metropolitan Achievement Test).

A clear definitive answer is not readily apparent for the first hypothesis, however, the data support rejection of the second hypothesis. A specific definition of student learning or student achievement is needed to interpret the findings of this study. If student achievement in reading is defined as the successful attainment of the instructional objectives that were developed for each reading course, then the CAM criterion measures are the more appropriate assessment of student learning. However, many educators still prefer to define and assess student achievement in more universal terms; i.e., standardized norm referenced achievement tests.

When employing the definition of student achievement as the successful attainment of the instructional objectives of the course, then the appropriate evidence to consider is the result of the CAM tests. The data indicate that the experimental groups which were provided CAM evaluation information at regular intervals throughout the course did somewhat better in reading achievement than the control groups. If a nationally normed standardized test is the definition of achievement then the results of the Metropolitan Achievement Test in reading is the appropriate criteria to be examined. When the reading achievement of the two groups is compared on the MAT, no clear direction is evidenced from the data. In one case (year 2, fifth grade) the experimental group had a statistically significant higher mean score on the MAT.

In another (year 3, fourth grade) the control group had a higher mean score. In most cases over the three years, however, there was no significant difference between the experimental and control groups on the MAT in reading. Thus, it could be concluded that the use of Comprehensive Achievement Monitoring has no major impact, either positively or negatively upon student achievement as defined by standardized norms.

This finding is not surprising since no attempt was made at the initiation of the project to increase reading scores of students on standardized tests. The intent was to increase student reading achievement in those skill areas defined by a set of instructional objectives developed specifically for the students involved in the West Seneca intermediate reading program. In this effort the CAM system seems successful.

It also can be concluded that the criterion referenced measures used in the CAM system are more sensitive indicators of student achievement than norm referenced tests. If educational decision makers want evaluation data on the effectiveness of their locally developed curricula, they are more likely to detect program strengths and weakness employing criterion referenced measures than standardized norm referenced tests. It should be noted that by the third year of the study the program objectives and criterion test items had been through two revisions. This fact along with the statistically significant differences between the experimental and control groups supports the contention that well developed criterion tests are more appropriate measures of the intended outcomes of an instructional program.

In all post-test comparisons between the student groups using CAM and those not using CAM, with the exception of the year 1, sixth grade group, the student achievement of reading skill objectives was significantly higher for those using CAM [5]. The data indicate that the experimental groups which were provided CAM evaluation information at regular intervals throughout the course increased their

[5] See tables in Section IV "Data Analysis" or in Appendix A for results obtained from the various tests.

reading skills to a greater extent than the control groups. The graphs depicting year end scores on CAM tests for both the experimental and control groups are found in Appendix B. In comparing the distribution of student scores it is evident that more students score at the upper ends of the scale in the CAM courses than in the control groups.

In examining class section comparisons (Tables for year one and for year three) it becomes evident that high achievement of course objectives is possible using a CAM system. Though it is possible to improve student achievement, simply installing a CAM evaluation system will not automatically guarantee increased achievement. Some of the sub-section mean scores in both years, one and three, attest to that.

Although the results of the CAM testing indicate significant differences in achievement levels between most of the experimental and control groups, the MAT results given at the same time to the same groups more often than not did not indicate a significant difference. The gain in achievement of the reading skill objectives was detected by the CAM tests while not necessarily by the MAT tests.

Since the higher achievement levels of the experimental groups was evidenced in all three years of the study and the fact that the fifth graders in year two (fourth graders, year one) maintained a higher achievement level after two years in the program, it would appear that the achievement increase is not due to the "Hawthorne Effect" alone. After three years of analyzing results, it appears that the CAM evaluation system does have a positive effect upon student achievement in reading, although the magnitude of the achievement increase is not large. One could conclude that the CAM system has the potential to significantly improve student achievement if utilized to its fullest.

When teachers involved in the experimental situation (designed and used the CAM system) were asked for their impressionistic evaluations and suggestions, one factor was emphasized. Teachers did not have enough time to examine in detail the analysis reports provided by the CAM system. They believed that if they had more time to jointly analyze the CAM test results, they could have greatly improved their instructional decision making. They would have been able to more adequately diagnose both individual and group learning problems and program weaknesses. The analysis of the data by teacher sub-sections supports the contention that when the data is understood and utilized by classroom teachers larger achievement gains result. Thus, the absence of dramatic differences in learning between the experimental and control groups after three years may be due to inadequate use of the analysis information provided by the CAM system. If teachers had spent more time reviewing the test results cooperatively with their colleagues and with the students, the effect upon student achievement may have been greater.

It would also appear that the use of a CAM evaluation system has some positive motivational impact upon students. The results of Question 6, Table 16, indicate a much larger percentage of parents of students in the experimental group believe their children's attitude toward reading has become more positive than those in the control group. A positive attitude toward reading may well have a long range effect upon increased achievement of reading skills.

Probably the clearest implication of the study is the need to replicate it on a much larger basis. Future studies should involve a broader sample of teachers and students and many different subject areas. Related variables such as teacher training, frequency of CAM testing and time allocated to data report analysis

ought to be examined. It may be that the potential benefit of a CAM system could be heightened substantially with a small incremental investment in in-service education and periodic released time for teachers.

APPENDIX A

TABLES OF TEST RESULT ANALYSIS

Table A-1

Comparison of Mean Scores on CAM Post tests - Grade 4 (year 1)

	Experimental Group	Control Group
Mean Score	16.41	13.31
Standard Deviation	3.50	3.60
Sample Population Size	155	143

Table A-2

Comparison of Mean Scores on CAM Tests - Grade 6 (year 1)

	Experimental Group	Control Group
Mean Score	10.84	10.75
Standard Deviation	2.80	2.99
Sample Population Size	128	88

Table A-3

Comparison of Mean Standard Schores on the Post MAT - Grade 4 (year 1)

	Experimental Group	Control Group
Mean Standard Score	74	71.2
Standard Deviation	11.62	12.85
Sample Population Size	154	139
Mean Grade Equivalent Score	4.64	4.28

Table A-4

Comparison of Mean Standard Scores on the MAT - Grade 6 (year 1)

	Experimental Group	Control Group
Mean Standard Score	88	90
Standard Deviation	13.15	14.07
Sample Population Size	131	88
Mean Grade Equivalent Score	6.8	7.0

Table A-5

Comparison of Gain Scores on the MAT - Grade 4 (year 1)

	Experimental Group	Control Group
Pre Mean Standard Score	65.9	63.0
Post Mean Standard Score	74.0	71.2
Gain Score	8.1	8.2

Table A-6

Comparison of Gain Scores on the MAT - Grade 6 (year 1)

	Experimental Group	Control Group
Pre Mean Standard Score	83.4	84.3
Post Mean Standard Score	88.0	90.0
Gain Score	4.6	5.7

Table A-7

Comparison of Mean Scores on CAM Tests - 5th grade (year 2)

	Experimental Group	Control Group
Mean Scores	25.2	23.1
Standard Deviation	5.1	5.9
Sample Population Size	140	129

Table A-8

Comparison of Mean Scores on MAT - 5th Grade (year 2)

	Experimental Group	Control Group
Mean Standard Score	83.0	80.9
Standard Deviation	12.1	12.0
Sample Population Size	145	128

Table A-9

Comparison of Pre-test Mean Scores on MAT - Grade 4 (year 3)

	Experimental Group	Control Group
Mean Standard Score	63.7	64.0
Mean Grade Equivalent Score	3.7	3.7
G.E. Standard Deviation	1.34	1.6
Sample Population Size	134	139

Table A-10

Comparison of Pre-test Mean Score on MAT - Grade 5 (year 3)

	Experimental Group	Control Group
Mean Standard Score	75.2	75.4
Mean Grade Equivalent Score	5.32	5.36
G.E. Standard Deviation	1.5	1.53
Sample Population Size	141	138

Table A-11

Mean Scores of Teacher Sub-sections for CAM tests - course 409 (year 3)

	Mean Score	Number of Students
Sub-section 1	12.3	4
Sub-section 2	19.8	11
Sub-section 3	20.1	17
Sub-section 4	20.3	10
Total Possible Score	24	

Table A-12

Mean Scores of Teacher Sub-section for CAM Tests - Course 419 (year 3)

	Mean Score	Number of Students
Sub-section 1	8.8	5
Sub-section 2	12.6	8
Sub-section 3	16.3	7
Sub-section 4	17.1	10
Sub-section 5	18.4	10
Sub-section 6	19.2	23
Sub-section 7	21.4	14
Sub-section 8	23.3	16
Sub-section 9	23.9	13
Sub-section 10	25.2	15
Sub-section 11	25.8	4
Total Possible Score	27	

Table A-13

Mean Scores of Teacher Sub-sections for CAM Tests - Course 509 (year 3)

	Mean Score	Number of Students
Sub-section 1	14.6	7
Sub-section 2	19.6	8
Sub-section 3	23.4	16
Sub-section 4	25.3	16
Sub-section 5	25.5	12
Sub-section 6	27.3	18
Sub-section 7	28.2	13
Sub-section 8	30.9	18
Total Possible Score	37	

APPENDIX B

Graphs of Student

CAM Test Scores

CHART 1

4th GRADE EXPERIMENTAL GROUP YEAR 1

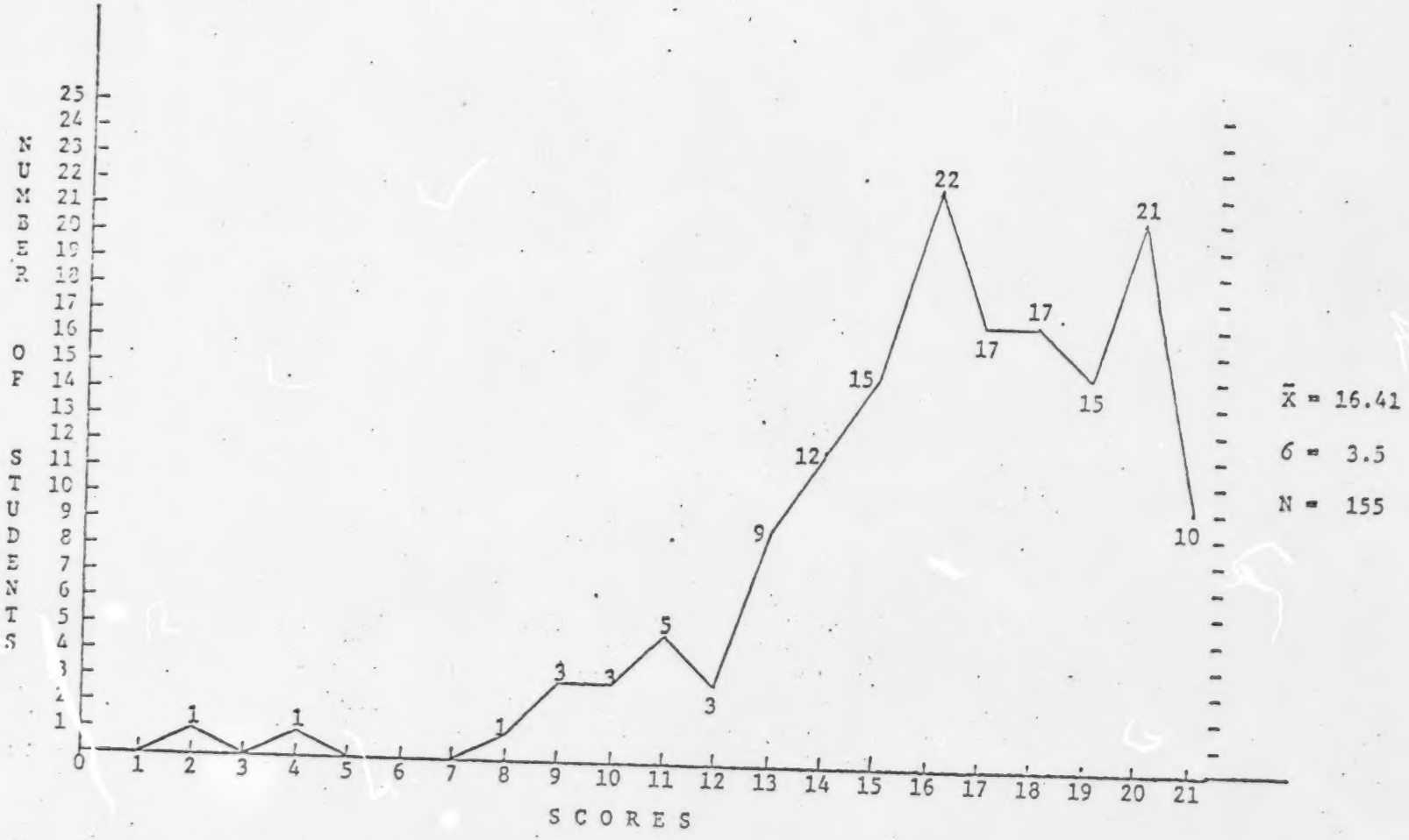


CHART 2

4th GRADE CONTROL GROUP YEAR 1

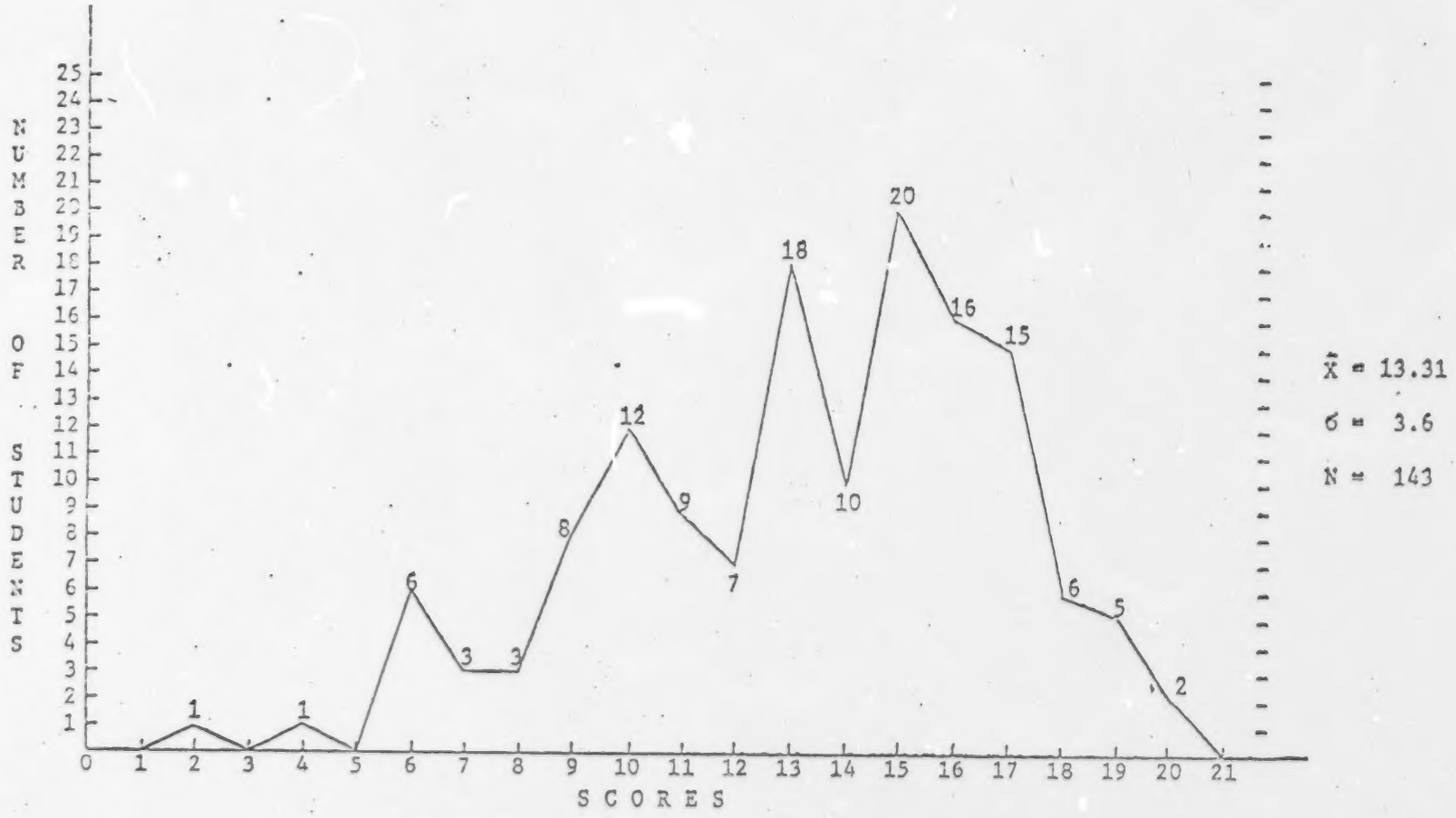


CHART 3

6th GRADE EXPERIMENTAL GROUP YEAR 1

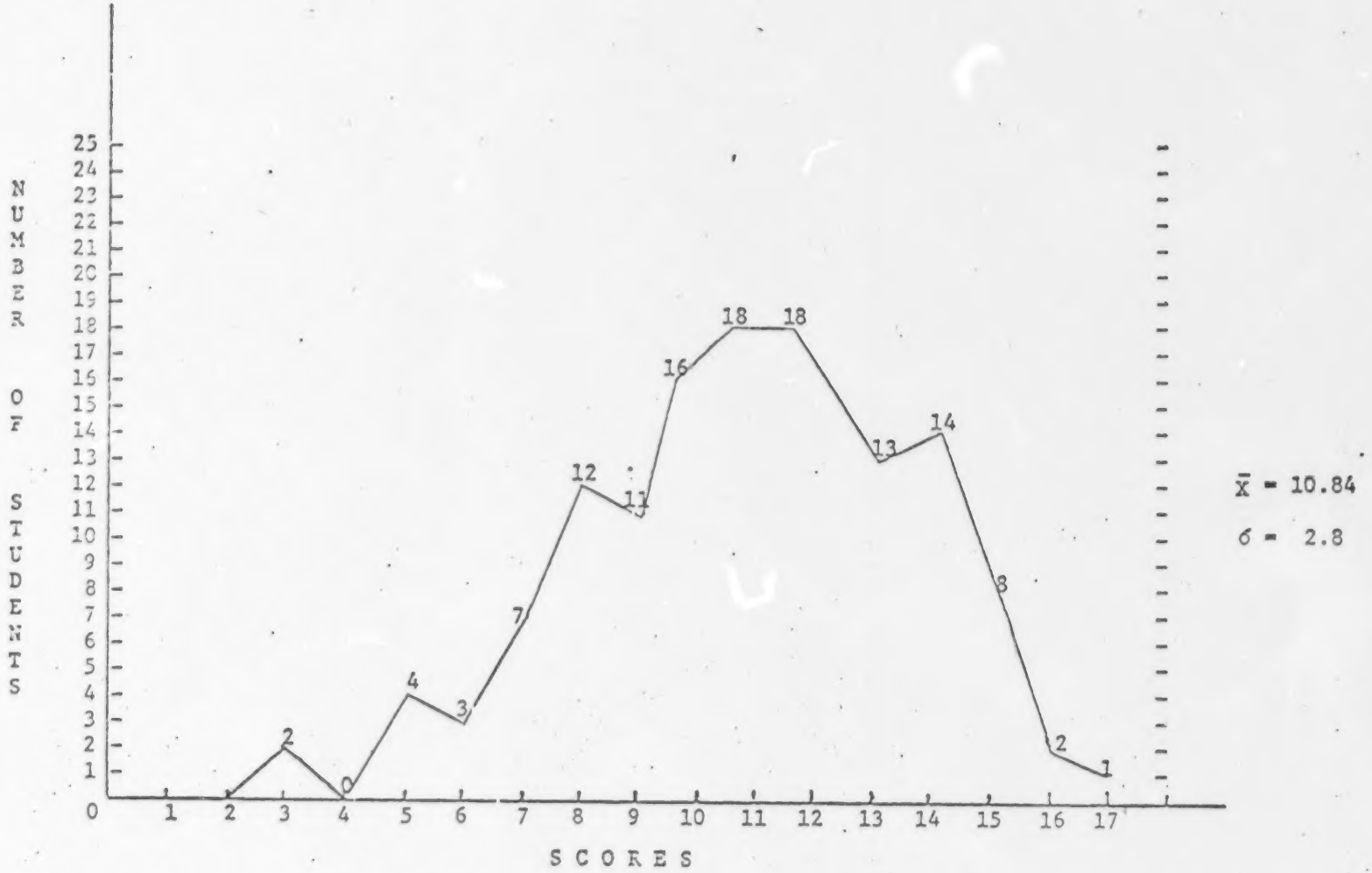


CHART 4

6th GRADE CONTROL GROUP YEAR 1

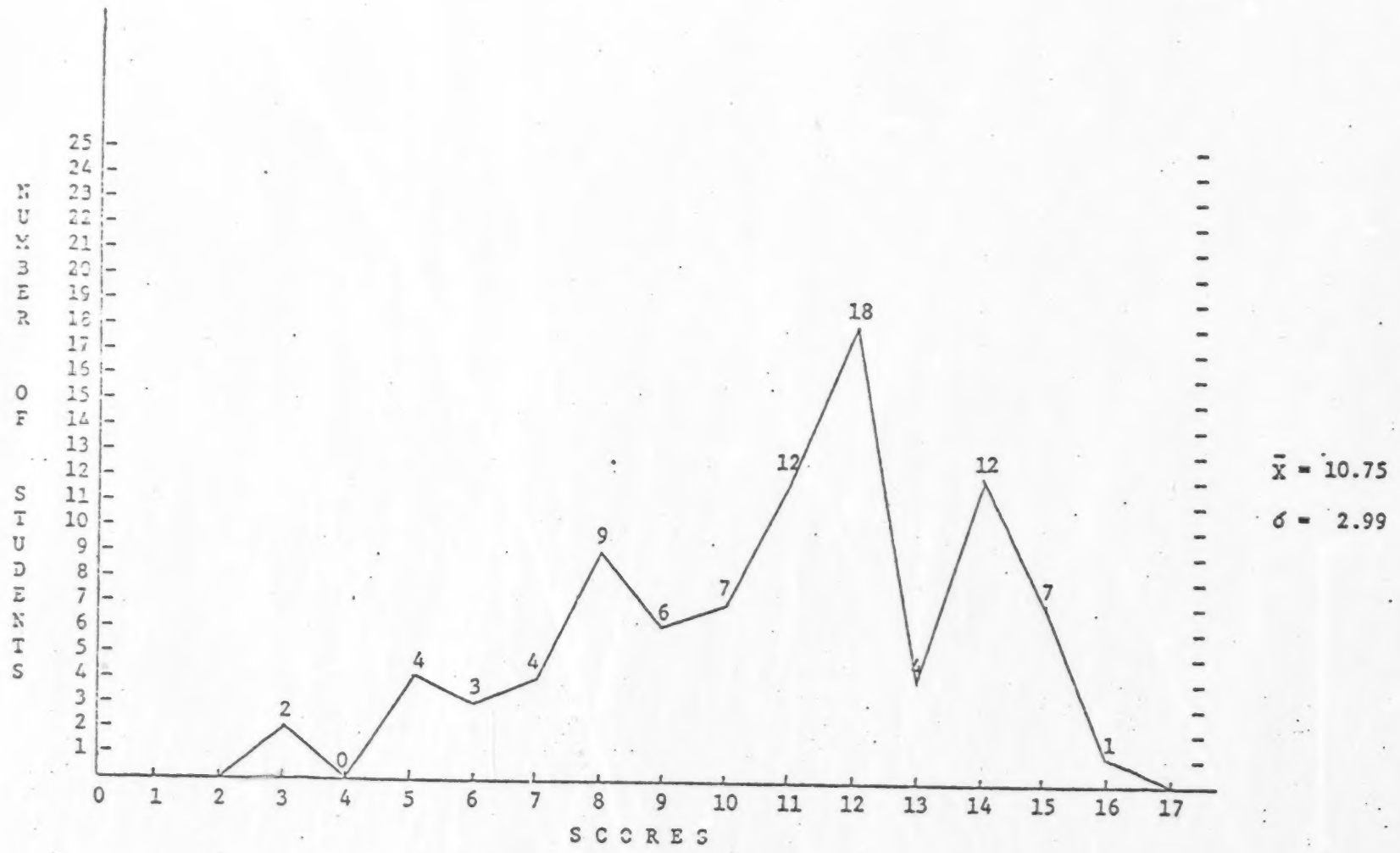


CHART 5

5th GRADE EXPERIMENTAL GROUP YEAR 2

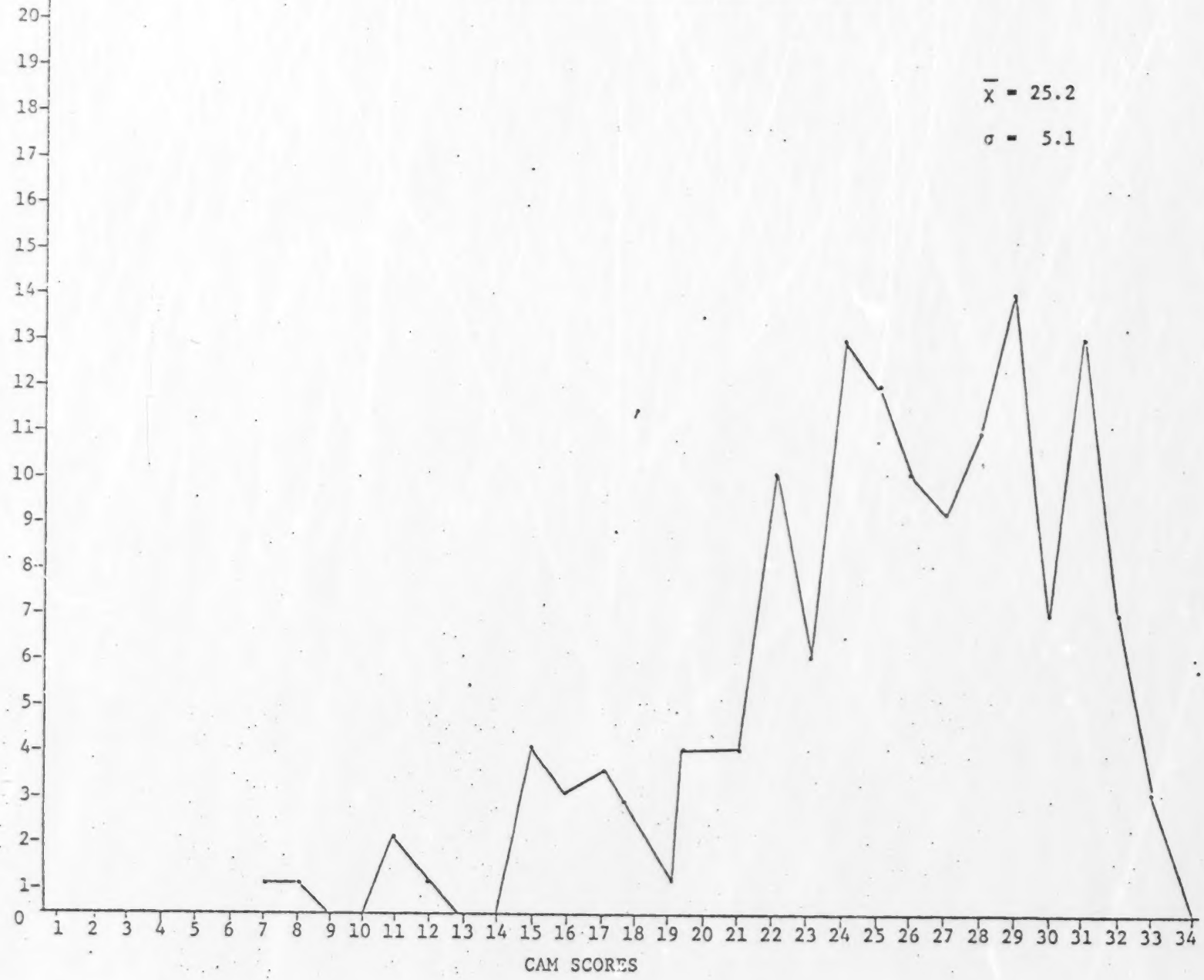


CHART 6

5th GRADE CONTROL GROUP YEAR 2

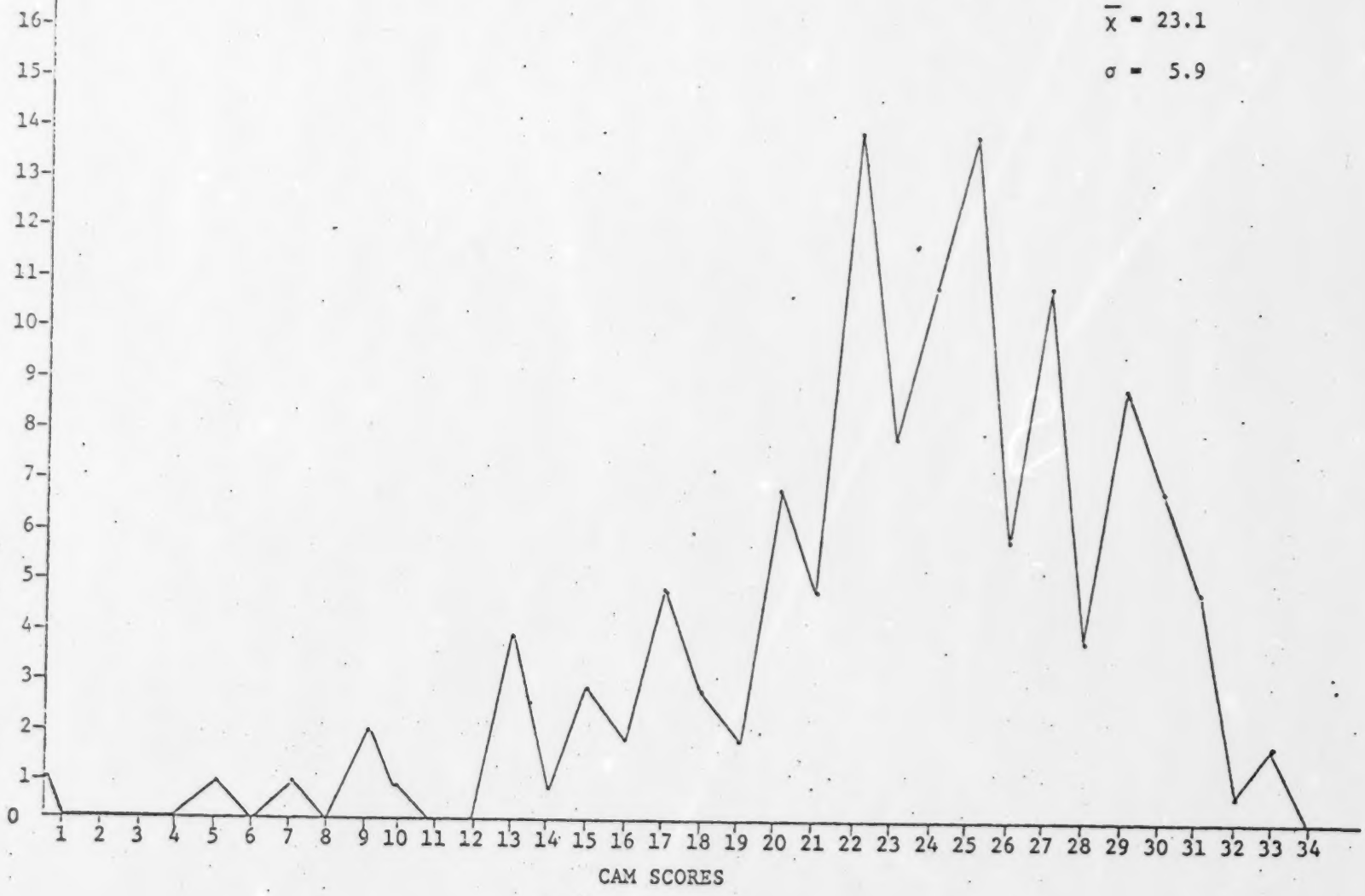
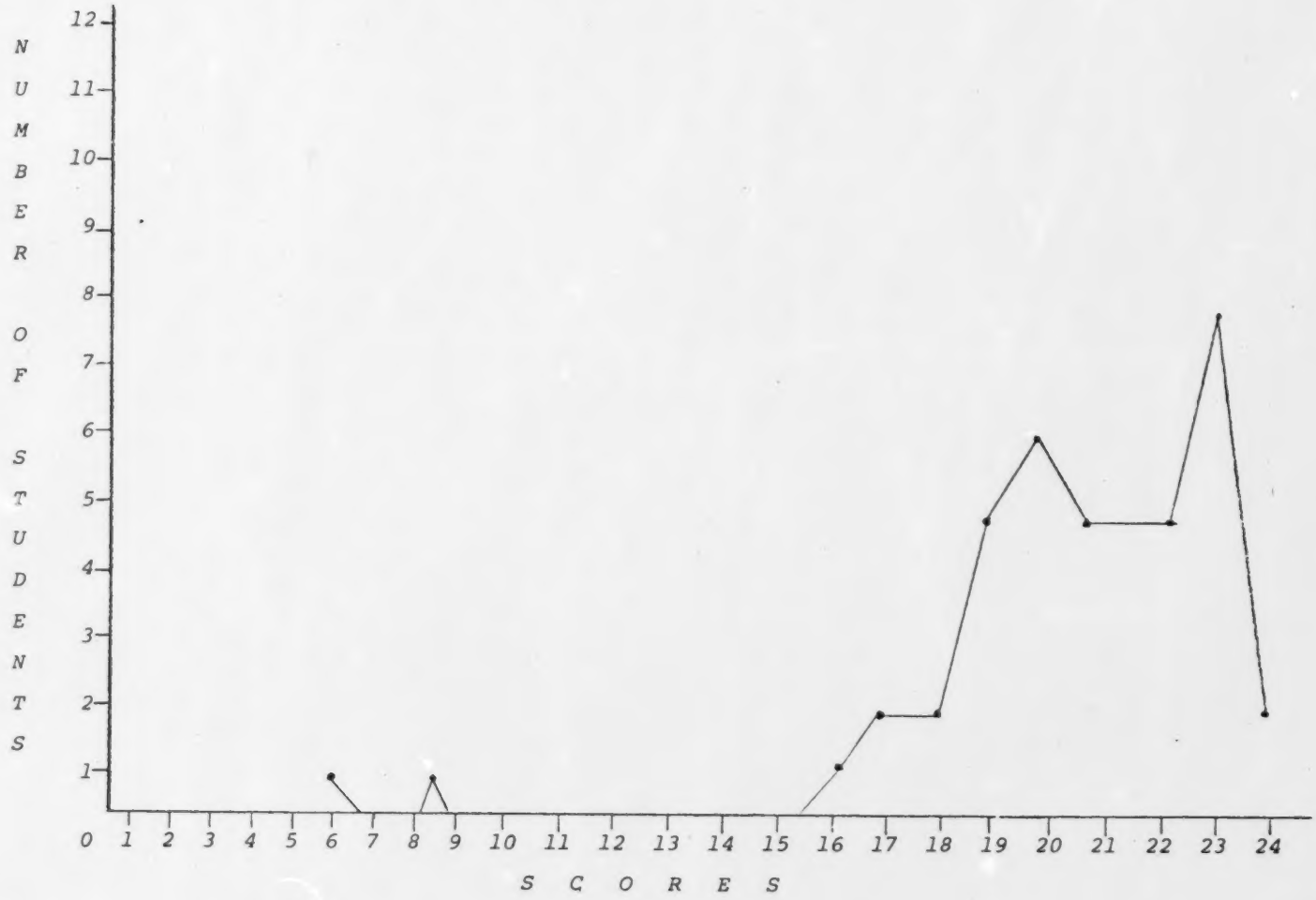


CHART 7

FOURTH GRADE (COURSE 409) EXPERIMENTAL GROUP - YEAR 3



Key: $\bar{X}$ = 19.3
σ = 3.7

CHART 8

FOURTH GRADE (COURSE 409) CONTROL GROUP - YEAR 3

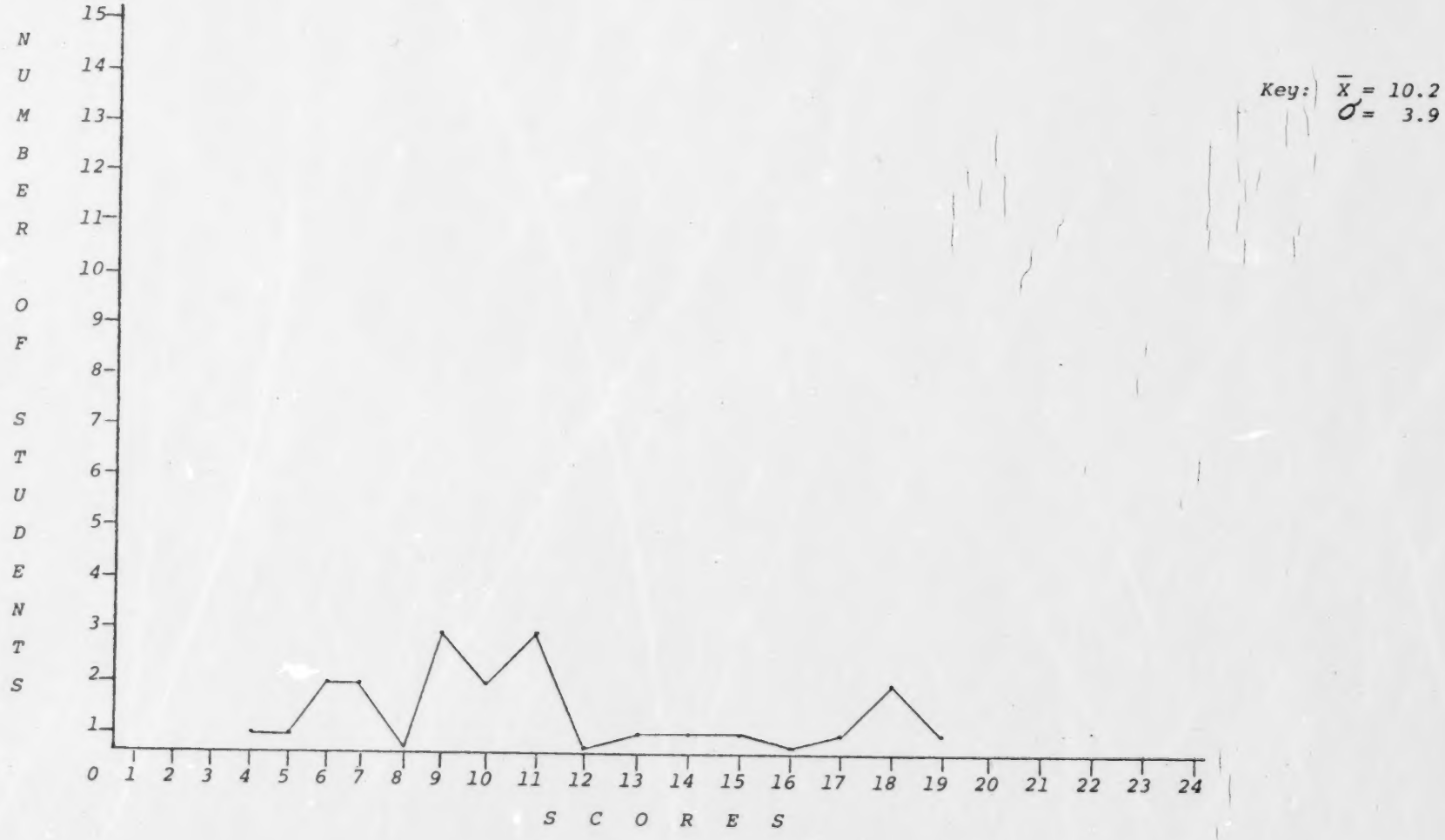


CHART 9

FOURTH GRADE (COURSE 419) EXPERIMENTAL GROUP - YEAR 3

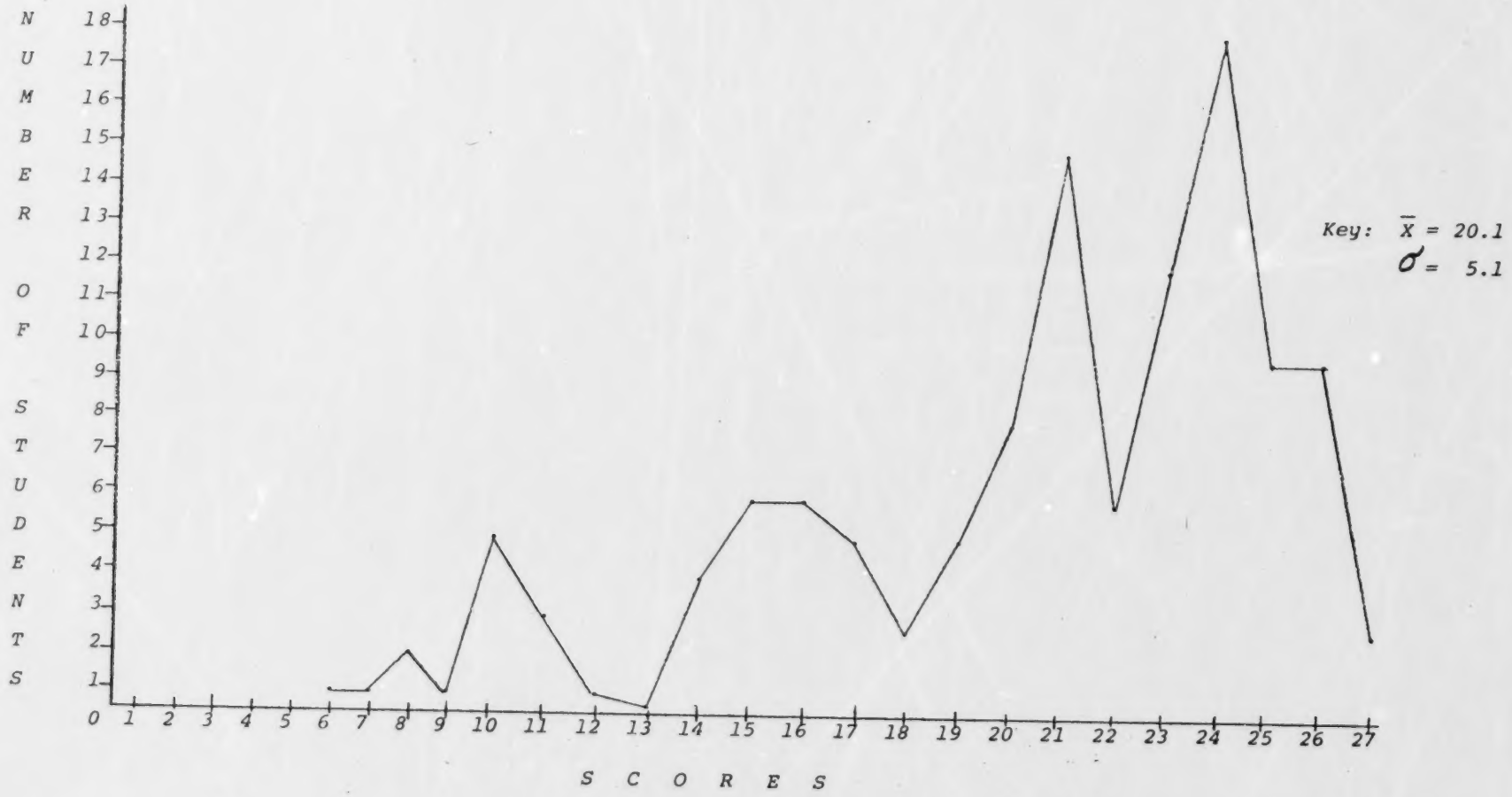
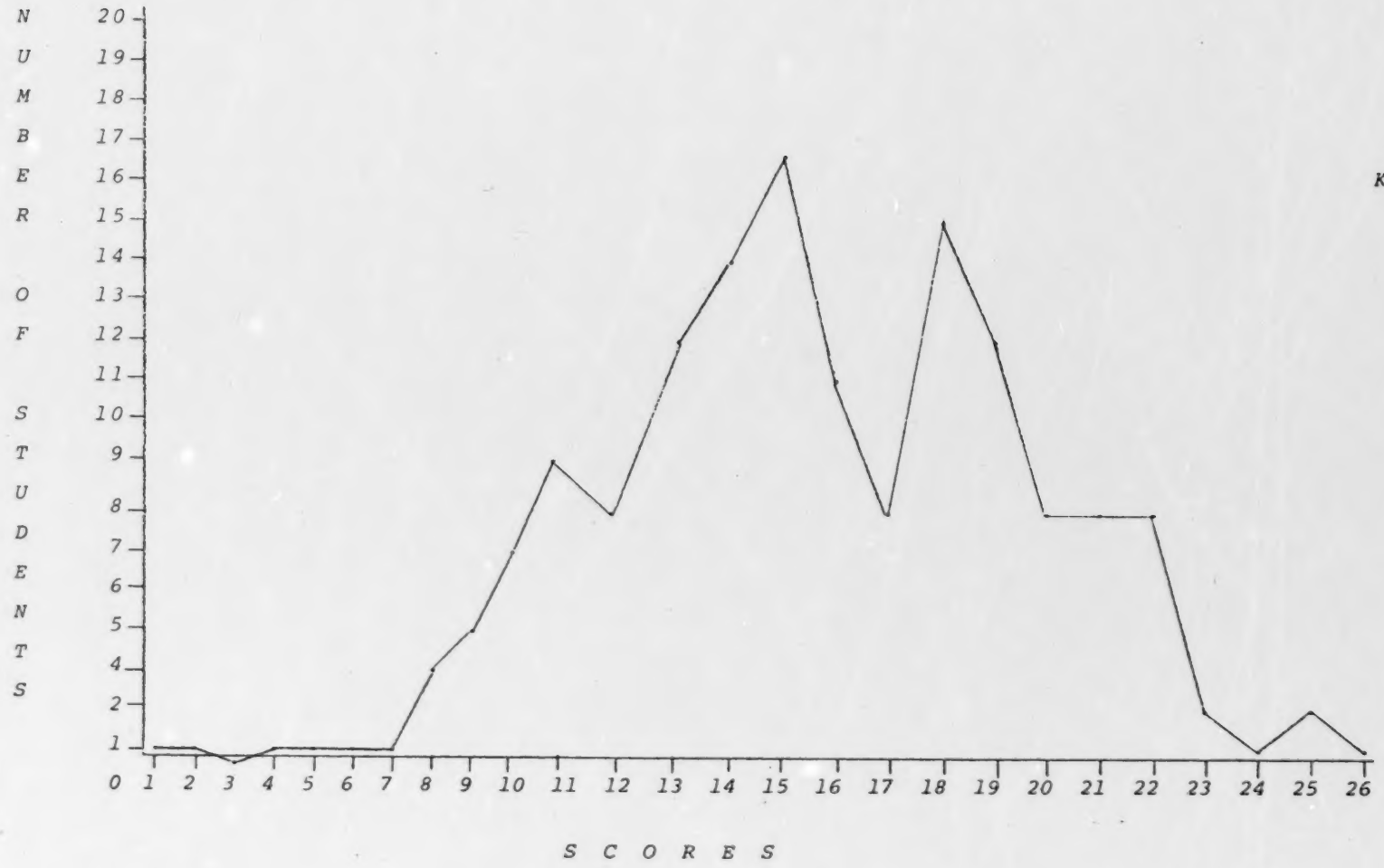


CHART 10

FOURTH GRADE (COURSE 419) CONTROL GROUP - YEAR 3



Key: $\bar{X}$ = 14.7

σ = 4.8

CHART 11

FIFTH GRADE (COURSE 509) EXPERIMENTAL GROUP - YEAR 3

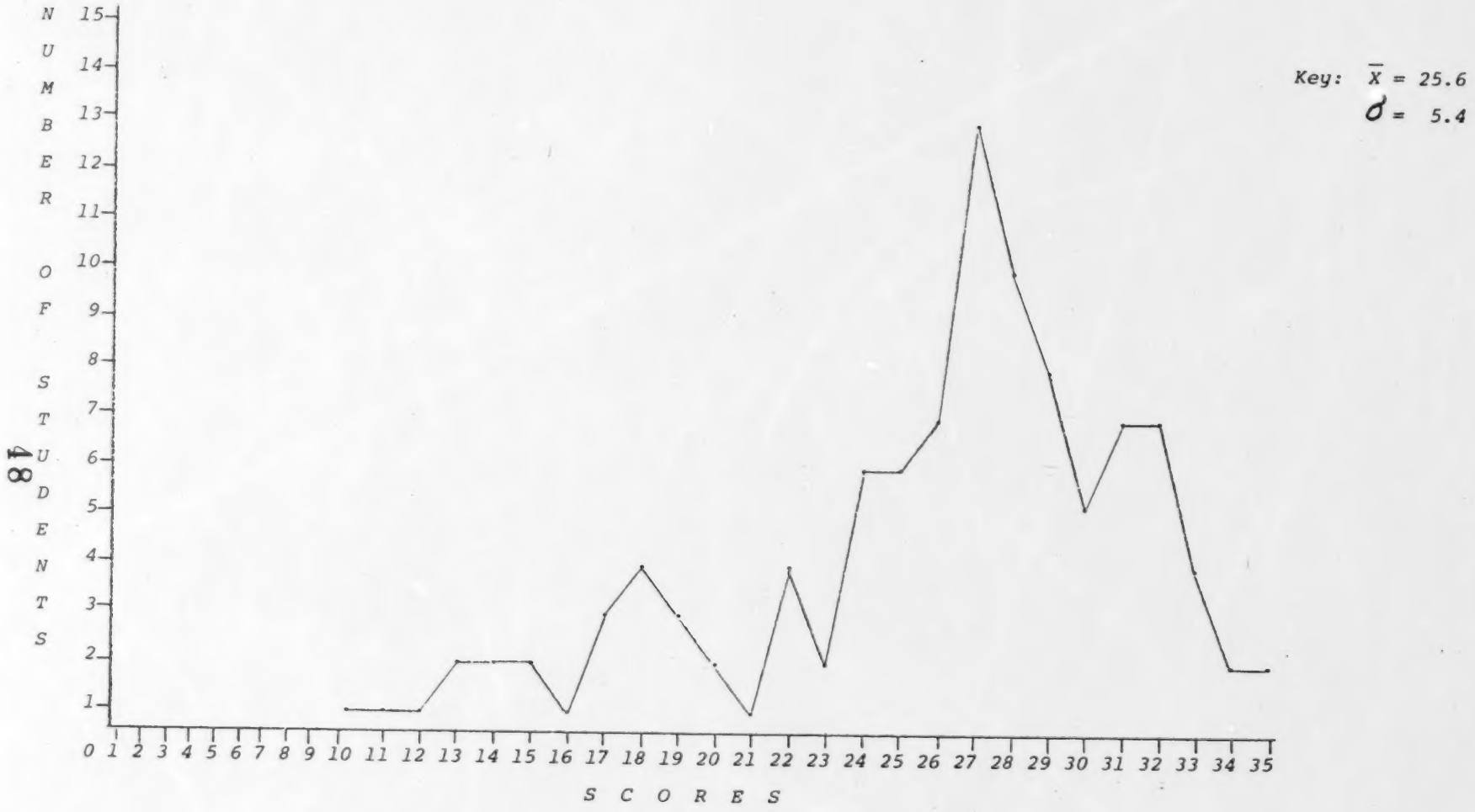
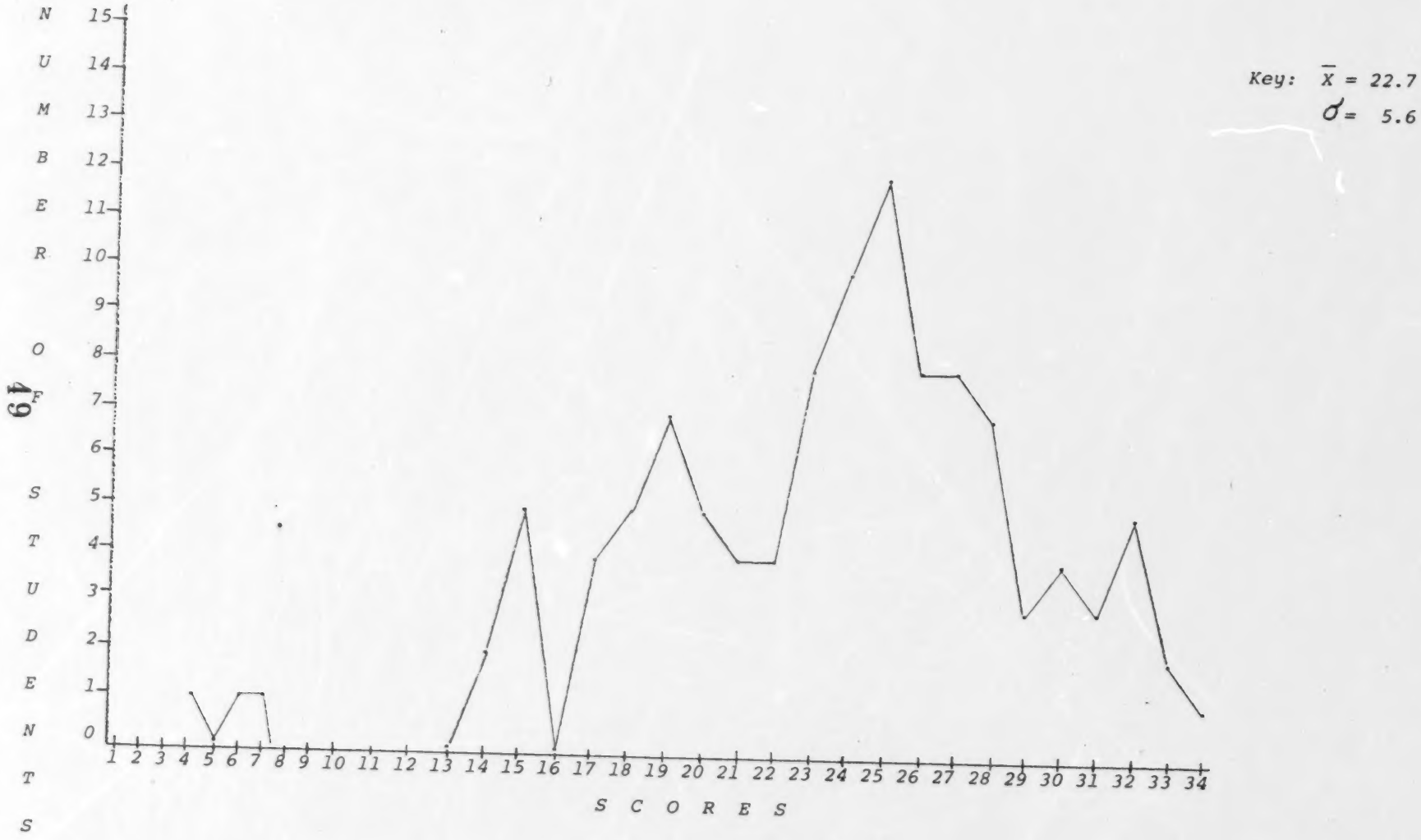


CHART 12

FIFTH GRADE (COURSE 509) CONTROL GROUP - YEAR 3



APPENDIX C

STUDENT BACKGROUND DATA QUESTIONNAIRE

_____ SCHOOL

_____ STREET

_____, NEW YORK

_____, 1974

Dear Parent:

Please provide information requested on the following form. This information is needed for your child's folder which follows him throughout his schooling.

This information is only for school records and is confidential.

_____
Principal

PUPIL'S LAST NAME _____ FIRST NAME _____ MIDDLE _____

BIRTHPLACE _____ DATE OF BIRTH _____ SEX _____

FATHER'S NAME _____ BIRTHPLACE _____

FATHER'S EDUCATION (a) ATTENDED COLLEGE ☐ Yes ☐ No Degree(s) earned _____

(b) GRADUATED FROM HIGH SCHOOL BUT DID NOT ATTEND COLLEGE _____

(c) ATTENDED BUT DID NOT GRADUATE FROM HIGH SCHOOL _____

(d) FINISHED 8TH GRADE BUT DID NOT ATTEND HIGH SCHOOL _____

(e) DID NOT FINISH 8TH GRADE _____

MOTHER'S NAME _____ BIRTHPLACE _____

MOTHER'S EDUCATION (a) ATTENDED COLLEGE _____

(b) GRADUATED FROM HIGH SCHOOL BUT DID NOT ATTEND COLLEGE _____

(c) ATTENDED BUT DID NOT GRADUATE FROM HIGH SCHOOL _____

(d) FINISHED 8TH GRADE BUT DID NOT ATTEND HIGH SCHOOL _____

(e) DID NOT FINISH 8TH GRADE _____

PUPIL'S PRESENT ADDRESS _____

PHONE _____

NOW LIVING WITH _____

FATHER'S OCCUPATION, WHERE EMPLOYED _____

MOTHER'S OCCUPATION, WHERE EMPLOYED _____

NAME OF BROTHERS _____ YEAR BORN _____

_____ YEAR BORN _____

_____ YEAR BORN _____

NAME OF SISTERS _____ YEAR BORN _____

_____ YEAR BORN _____

_____ YEAR BORN _____

HAS THE PUPIL WHOSE NAME APPEARS ON THE FRONT PAGE ATTENDED SCHOOLS OTHER THAN THE PRESENT ONE _____ YES ☐ NO ☐

IF YES, HOW MANY OUTSIDE OF WEST SENECA _____

IN YOUR JUDGEMENT, HAS YOUR CHILD'S ATTITUDE ABOUT READING INSTRUCTION CHANGED POSITIVELY IN THE LAST TWO YEARS? YES _____ NO _____ NO CHANGE _____

DOES YOUR CHILD MAKE USE OF THE LIBRARY SERVICES MADE AVAILABLE?

Check those which apply

_____ School library
_____ Bookmobile
_____ Public library

IN YOUR JUDGEMENT, HOW WOULD YOU RATE THE CHILD'S GENERAL HEALTH?

_____ Poor
_____ Fair
_____ Good
_____ Excellent

IT IS VERY IMPORTANT THAT YOU RETURN THIS TO YOUR CHILD'S HOMEROOM TEACHER AS SOON AS POSSIBLE.